

**APPENDIX 1: CLEAN COPY OF PENDING CLAIMS OF U.S. SERIAL NO. 09/684,066**

1. A method of identifying one or more positions in a polymer family, the method comprising:
  - (a) accessing data representing a multiple sequence alignment (MSA) of a plurality of polymer sequences; and
  - (b) identifying one or more evolutionarily conserved amino acid positions within the MSA using the following equation:

$$\Delta G_i^{stat} = kT^* \sqrt{\sum_x \left( \ln \frac{P_i^x}{P_{MSA}^x} \right)^2}$$

wherein:

- i is a position in the MSA;
- $\Delta G_i^{stat}$  is the conservation energy value for position i;
- $P_i^x$  is the probability of monomer x at position i;
- $P_{MSA}^x$  is the probability of monomer x in the MSA; and
- kT\* is an energy unit, where k is Boltzmann's constant.

2. The method of claim 1, wherein the method is executed using a machine.
3. A program storage device readable by the machine of claim 2 and encoding instructions executable by the machine for performing the operations recited in the claim.
4. The method of claim 1, further comprising generating a graphical image of the conservation energy values.
5. The method of claim 1, wherein the polymer sequences comprise protein sequences.
6. The method of claim 1, wherein monomer x comprises amino acid x.

7. The method of claim 1, wherein the data accessed comprises data from the PSD-95 (Postsynaptic density protein of Mr 95kDa), Dlg (Drosophila Discs-Large protein) and ZO-1 (Zonula occludens protein 1) protein family.
8. The method of claim 1, wherein the data accessed comprises data from the p21<sup>ras</sup> protein family.
9. The method of claim 1, wherein the data accessed comprises data from the hemoglobin protein family.
10. A method of identifying one or more positions in a polymer family, the method comprising:
  - (a) accessing data representing a multiple sequence alignment (MSA) of a plurality of polymer sequences;
  - (b) calculating a conservation energy value for each position in the MSA using the following equation:

$$\Delta G_i^{stat} = kT^* \sqrt{\sum_x \left( \ln \frac{P_i^x}{P_{MSA}^x} \right)^2}$$

wherein:

- i is a position in the MSA;
- $\Delta G_i^{stat}$  is the conservation energy value for position i;
- $P_i^x$  is the probability of monomer x at position i;
- $P_{MSA}^x$  is the probability of monomer x in the MSA;
- kT\* is an energy unit, where k is Boltzmann's constant; and
- (c) identifying one or more positions within the MSA that have statistically significant conservation energy values.

11. The method of claim 10, wherein the method is executed using a machine.

12. A program storage device readable by the machine of claim 11 and encoding instructions executable by the machine for performing the operations recited in the claim.
13. The method of claim 10, further comprising generating a graphical image of the conservation energy values.
14. The method of claim 10, wherein the polymer sequences comprise protein sequences.
15. The method of claim 10, wherein monomer x comprises amino acid x.
16. The method of claim 10, wherein the data accessed comprises data from the PSD-95 (Postsynaptic density protein of Mr 95kDa),Dlg (Drosophila Discs-Large protein) and ZO-1 (Zonula occludens protein 1) protein family.
17. The method of claim 10, wherein the data accessed comprises data from the p21<sup>ras</sup> protein family.
18. The method of claim 10, wherein the data accessed comprises data from the hemoglobin protein family.